# The 1.7 angstrom crystal structure of the apo form of the soluble quinoprotein glucose dehydrogenase from Acinetobacter calcoaceticus reveals a novel internal conserved sequence repeat

Oubrie, A; Rozeboom, HJ; Kalk, KH; Duine, JA; Dijkstra, BW

*Published in:*
Journal of Molecular Biology

Article No. jmbi.1999.2766 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 289, 319–333

JMB

# The 1.7 Å Crystal Structure of the Apo Form of the Soluble Quinoprotein Glucose Dehydrogenase from *Acinetobacter calcoaceticus* Reveals a Novel Internal Conserved Sequence Repeat

**Arthur Oubrie[1], Henriëtte J. Rozeboom[1], Kor H. Kalk[1] Johannis A. Duine[2] and Bauke W. Dijkstra[1]***

[1]*Laboratory of Biophysical Chemistry and BIOSON Research Institute, University of Groningen, Nijenborgh 4 9747 AG Groningen The Netherlands*

[2]*Department of Microbiology and Enzymology, Delft University of Technology Julianalaan 67, 2628 BC Delft The Netherlands*

**Corresponding author*

The crystal structure of a dimeric apo form of the soluble quinoprotein glucose dehydrogenase (s-GDH) from *Acinetobacter calcoaceticus* has been solved by multiple isomorphous replacement followed by density modification, and was subsequently refined at 1.72 Å resolution to a final crystallographic *R*-factor of 16.5 % and free *R*-factor of 0.8 %. The s-GDH monomer has a β-propeller fold consisting of six four-stranded anti-parallel β-sheets aligned around a pseudo 6-fold symmetry axis. The enzyme binds three calcium ions per monomer, two of which are located in the dimer interface. The third is bound in the putative active site, where it may bind and functionalize the pyrroloquinoline quinone (PQQ) cofactor. A data base search unexpectedly showed that four uncharacterized protein sequences are homologous to s-GDH with many residues in the putative active site absolutely conserved. This indicates that these homologs may have a similar structure and that they may catalyze similar PQQ-dependent reactions.

A structure-based sequence alignment of the six four-stranded β-sheets in s-GDH's β-propeller fold shows an internally conserved sequence repeat that gives rise to two distinct conserved structural motifs. The first structural motif is found at the corner of the short β-turn between the inner two β-strands of the β-sheets, where an Asp side-chain points back into the β-sheet to form a hydrogen-bond with the OH/NH of a Tyr/Trp side-chain in the same β-sheet. The second motif involves an Arg/Lys side-chain in the C β-strand of one β-sheet, which forms a bidentate salt-bridge with an Asp/Glu in the CD loop of the next β-sheet. These intra and inter-β-sheet hydrogen-bonds are likely to contribute to the stability of the s-GDH β-propeller fold.



*Keywords:* electron transfer; glucose dehydrogenase; quinoprotein; PQQ; X-ray structure

## Introduction

Quinoproteins constitute a class of enzymes that use quinone cofactors for catalysis (Duine, 1991). The quinoprotein glucose dehydrogenase (GDH) uses pyrroloquinoline quinone (PQQ) as a cofactor. Two different types of PQQ-dependent GDHs have been found: a membrane-bound form (m-GDH), which is widespread in Gram-negative bacteria, and a soluble type (s-GDH), which has so far exclusively been identified in the periplasmic space of the bacterium *Acinetobacter calcoaceticus*

Abbreviations used: GDH, quinoprotein glucose dehydrogenase; m-GDH, membrane-bound GDH; s-GDH, soluble GDH; MDH, quinoprotein methanol dehydrogenase; PQQ, pyrroloquinoline quinone; NCS, non-crystallographic symmetry; PEG, polyethylene glycol; $F_o$, observed structure factor; $F_c$, calculated structure factor; rms, root-mean-square; $R_{crys}$, crystallographic *R*-factor; $R_{free}$, free *R*-factor.

E-mail address of the corresponding author: bauke@chem.rug.nl

0022-2836/99/220319–15 $30.00/0

(Cleton-Jansen *et al.*, 1988; Matsushita *et al.*, 1989). The two types of GDH have no sequence homology (Cleton-Jansen *et al.*, 1990), and are kinetically and immunologically distinct (Matsushita *et al.*, 1989).

m-GDH and some other PQQ-dependent proteins show significant sequence identity with the α-subunit of the PQQ-dependent methanol dehydrogenases (MDH; Cleton-Jansen *et al.*, 1990). All these enzymes are consequently expected to have a fold similar to MDH, the only PQQ-dependent protein with known three-dimensional structure (Ghosh *et al.*, 1995; Xia *et al.*, 1992, 1996). By contrast, the sequence of s-GDH is unique among the known PQQ-dependent protein sequences (Cleton-Jansen *et al.*, 1988).

s-GDH is a dimer of identical subunits of 50 kDa (454 residues) each (Dokter *et al.*, 1986). One dimer has been reported to bind two PQQ molecules and four calcium ions (Geiger & Görisch, 1989; Olsthoorn *et al.*, 1997). $Ca^{2+}$ has a dual role in s-GDH, being required for dimerization as well as for functionalization of the bound PQQ (Olsthoorn & Duine, 1996; Olsthoorn *et al.*, 1997). The enzyme converts a wide range of pentose and hexose sugars, mono- as well as disaccharides, into their corresponding lactones (Matsushita *et al.*, 1989). s-GDH is able to donate electrons to several neutral or cationic, artificial electron acceptors, including *N*-methylphenazonium methyl sulphate (Olsthoorn & Duine, 1996) and electroconducting polymers (Ye *et al.*, 1993). The physiological electron acceptor of s-GDH is unknown, although electrons can be transferred from s-GDH to a soluble cytochrome *b*562 (Dokter *et al.*, 1988).

s-GDH is a stable enzyme with a high turnover number and (electrochemical) reoxidation of the PQQ cofactor can be easily achieved without negative effects of $O_2$ tension and/or product inhibition. Therefore, the protein has been applied in a wired biosensor (Ye *et al.*, 1993) and on test strips to monitor glucose concentrations in blood samples of diabetes patients (Hoenes, 1993, 1994; Hoenes & Unkrig, 1996). Engineering experiments, to confine the substrate specificity to glucose and thus to make glucose detection more specific, have awaited the crystal structure determination of this protein. Here, we describe the crystallization, structure determination and refinement to 1.72 Å of an apo form of recombinant s-GDH. The model confirms that s-GDH is a dimeric molecule with a water-mediated dimer interface. The dimeric molecule contains a total of six calcium ions, four of which are located close to the dimer interface. Two other calcium ions are present in the putative active sites of the two domains. Each monomer is folded as a β-propeller, consisting of six four-stranded anti-parallel β-sheets. A structure-based sequence alignment of these β-sheets shows an internally conserved sequence repeat that gives rise to two distinct conserved structural motifs.

## Results and Discussion

### Crystallization

s-GDH has proven a difficult protein to crystallize. Initial crystallization experiments, using the hanging drop vapor diffusion method at room temperature, were carried out by Geiger & Görisch (1986), who reported thin clusters of crystals unsuitable for X-ray diffraction analysis. Schlunegger *et al.* (1993) obtained s-GDH crystals that diffracted to beyond 2.1 Å with a slightly different procedure. A crystal structure, however, could not be determined. This has been attributed to a combination of strong non-isomorphism (even between native crystals), and the lack of a suitable stabilization buffer (M. Grütter, personal communication). In order to grow stable, well-diffracting crystals, we have optimized the procedure used by Schlunegger *et al.* (1993). It was found necessary to purify the enzyme immediately prior to crystallization. Protein material prepared in this way could be crystallized reproducibly. A newly developed artificial mother liquor further improved the stability and diffraction quality of the crystals (see below).

### Structure determination

As no homologous quinoprotein is known (Cleton-Jansen *et al.*, 1988), the three-dimensional structure of s-GDH was solved by the multiple isomorphous replacement method followed by density modification. An extensive heavy-atom search resulted in the identification of four heavy-atom derivatives (Table 1). Both Harker and cross-peaks in the $K_2PtCl_4$ difference Patterson map could be fully interpreted, yielding two unique $Pt^{2+}$-sites. Cross-difference Fourier maps calculated for the other derivatives, using phases derived from the platinum sites, identified a total of ten heavy-atom sites in the four derivative data sets (Table 1). After heavy-atom parameter and phase refinement, an overall figure of merit of 0.64 was obtained (Table 1). A 3.0 Å electron density map, obtained after solvent flattening, allowed the unambiguous tracing of approximately 300 out of 454 $C^{\alpha}$ positions per monomer. No interpretable density could be observed for the remainder of the molecule. At this point, the non-crystallographic symmetry-averaging procedure of the program DM (Cowtan & Main, 1993), was used to improve the solvent-flattened phases. With a bootstrapping procedure, described earlier by Rudenko *et al.* (1996), $C^{\alpha}$ positions for 436 of the 454 residues of one monomer could be obtained. The amino acid sequence was fitted to the electron density and an initial model was built using the program O (Jones *et al.*, 1991).

### Refinement and quality of the model

Final crystallographic ($R_{crys}$) and free ($R_{free}$: Brünger, 1992) *R*-factor values of 16.5 % and 0.8 %,

**Table 1.** Data collection, heavy-atom parameter and phasing statistics

| | Native | $K_2PtCl_4$ I | $K_2PtCl_4$ II | $KAuBr_4$ I | $KAuBr_4$ II | $CdCl_2$ | $K_2OsO_4$ |
|---|---|---|---|---|---|---|---|
| X-ray source | Rotating anode | Rotating anode | X-11 | Rotating anode | X-31 | Rotating anode | Rotating anode |
| Temperature (K) | 293 | 293 | 100 | 293 | 293 | 293 | 293 |
| Resolution (Å) | 2.7 | 2.5 | 1.7 | 2.8 | 2.5 | 3.0 | 2.5 |
| No. of unique reflections | 25,355 | 32,656 | 83,663 | 23,950 | 31,728 | 19,069 | 30,728 |
| Completeness (%) | 96.9 | 97.9 | 85.9 | 97.5 | 96.6 | 99.2 | 92.8 |
| In last shell (%) | 84.5 | 88.3 | 27.3 | 52.0 | 94.0 | 98.3 | 58.1 |
| $R_{merge}$ $(I)$[a] (%) | 6.3 | 6.0 | 5.8 | 5.1 | 7.7 | 6.6 | 8.4 |
| No. of heavy-atom sites | 2 | 2 | 2 | 2 | 4 | 2 | |
| Phasing power | | 0.66 | | | 1.74 | 1.06 | 0.33 |
| Overall mean figure of merit[b] for phases to 3.0 Å: 0.64 | | | | | | | |

[a] $R_{merge}(I)$: $\Sigma_{hkl}\Sigma_i\,|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i\,I_i(hkl)$.
[b] Figure of merit $= |\Sigma_\alpha P(\alpha)F_{hkl}(\alpha)/\Sigma_\alpha P(\alpha)|/|F_{hkl}|$.

respectively, were obtained for data from 20.0 to 1.72 Å resolution (no σ cut-off). The model contains two s-GDH monomers, six calcium ions, two platinum ions and 1233 water molecules. Of the 454 residues making up a complete monomer, residues 105-110, 335-344 in both protein molecules and the C-terminal two and four residues for monomers A and B, respectively, have been omitted from the final model because no electron density could be observed for these residues. The remainder of the protein residues are very well defined in the electron density maps (e.g. see Figure 1). The overall geometry of the model is good, as can be deduced from the root-mean-square (rms) differences from ideal geometry of 0.011 Å and 2.3° for bond lengths and angles, respectively (Engh & Huber, 1991). A Ramachandran plot shows that 87.6% and 11.8% of the residues have most favored and additional allowed combinations of ϕ/ψ-angles (not shown: Ramakrishnan & Ramachandran, 1965).

There are three outliers in each monomer: Leu169 and Ser146 are found in a generously allowed region and a disallowed region, respectively, and Pro248 is found in a region that is highly unusual for proline. Leu169 is located in a loop at the surface of the protein and has well-defined electron density. Ser146 is located in one of the β-strands that make up the core of the protein. Electron density for this residue is very convincing (Figure 1) and attempts to impose more regular ϕ/ψ-angles onto this residue resulted in the vanishing of electron density. Furthermore, the hydroxyl group and peptide backbone of Ser146 make numerous hydrogen-bonds, thereby stabilizing its energetically unfavorable conformation. A structural or functional role for this residue, however, is not obvious. In contrast, the peptide backbone carbonyl groups around Pro248, for which electron density is excellent, are ligands of one calcium ion, which is located at the top of the superbarrel, close to the 6-fold pseudo-symmetry axis.

Pro266 and Tyr325 have a *cis*-peptide conformation. The cyclic ring of Pro266 makes a hydrophobic stacking interaction with Trp265, and Tyr325 is positioned in a sharp turn and makes a

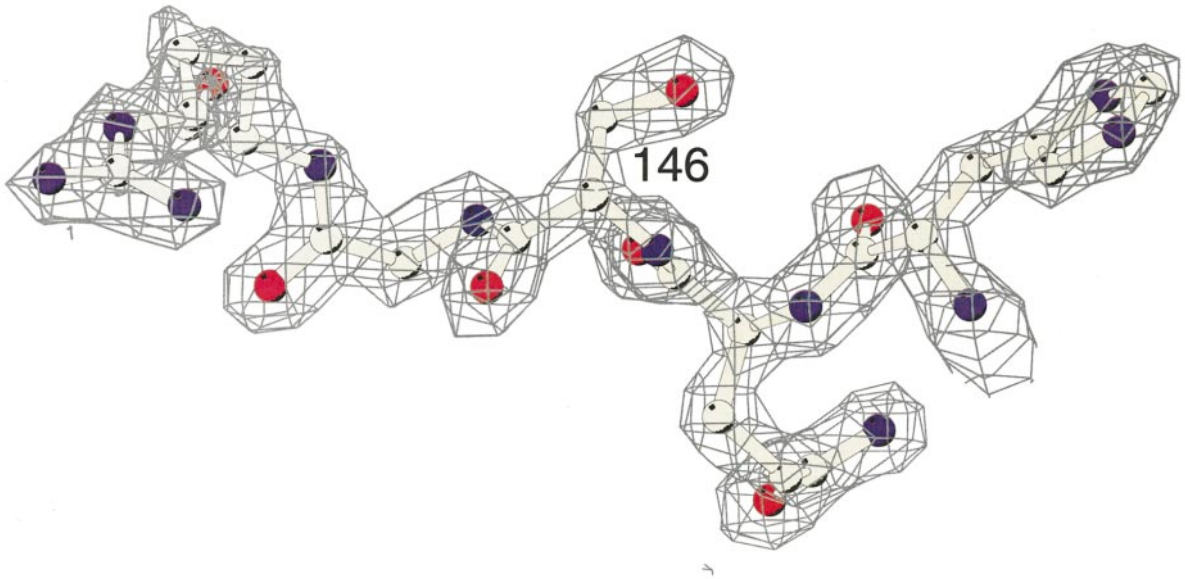


**Figure 1.** Representative part of the final $\sigma_A$-weighted (Read, 1986) $2F_o - F_c$ electron density map, contoured at 1.2 σ and centered around Ser146, which is the only residue with ''unallowed'' ϕ/ψ combinations.

**Table 2.** Statistics of structure refinement and quality of the s-GDH model

| | |
|---|---|
| Resolution range (Å) | 20.0-1.72 |
| Completeness (%) (test set) | 81.6 (4.3) |
| Completeness 1.80-1.72 Å (%) (test set) | 69.9 (3.7) |
| No. of reflections (test set) | 77,210 (4102) |
| $R_{crys}$ [a] (%) | 16.5 |
| $R_{free}$ [b] (%) | 20.8 |
| No. of non-H atoms in protein model | 6815 |
| No. of water molecules | 1233 |
| rms deviation bond lengths (Å) | 0.011 |
| rms deviation bond angles (°) | 2.3 |

[a] $R_{crys} = \Sigma_{hkl} | |F_{obs,\ work}| - k|F_{calc}| | / \Sigma_{hkl}|F_{obs,\ work}| \times 100$.
[b] $R_{free} = \Sigma_{hkl} | |F_{obs,\ test}| - k|F_{calc}| | / \Sigma_{hkl}|F_{obs,\ test}| \times 100$.

hydrophobic stacking interaction with Leu324. A similar hydrophobic interaction has been identified in ten out of 43 non-proline *cis*-peptide bonds (Jabs *et al.*, 1999). The values of bond lengths and angles of the two non-proline *cis*-peptide bonds lie within 1 σ of the average geometrical parameters published by Jabs *et al.* (1999). Statistical details of the structure refinement and the quality of the model are summarized in Table 2.

## Overall structure

The overall structure of the s-GDH dimer is shown in Figure 2. The dimer has approximate dimensions of 100 Å × 50 Å × 50 Å. One protein monomer consists of one, roughly globular domain with a radius of 50 Å. The two monomers in the dimer interact across a considerable, planar interface (see below).

## Structure of the s-GDH monomer

The building block of the protein core of one s-GDH monomer is a four-stranded anti-parallel β-sheet, which is repeated six times in the structure (Figure 3). The pronounced right-handed twist present in all six β-sheets gives them the appearance of a propeller-blade. All six β-sheets are aligned

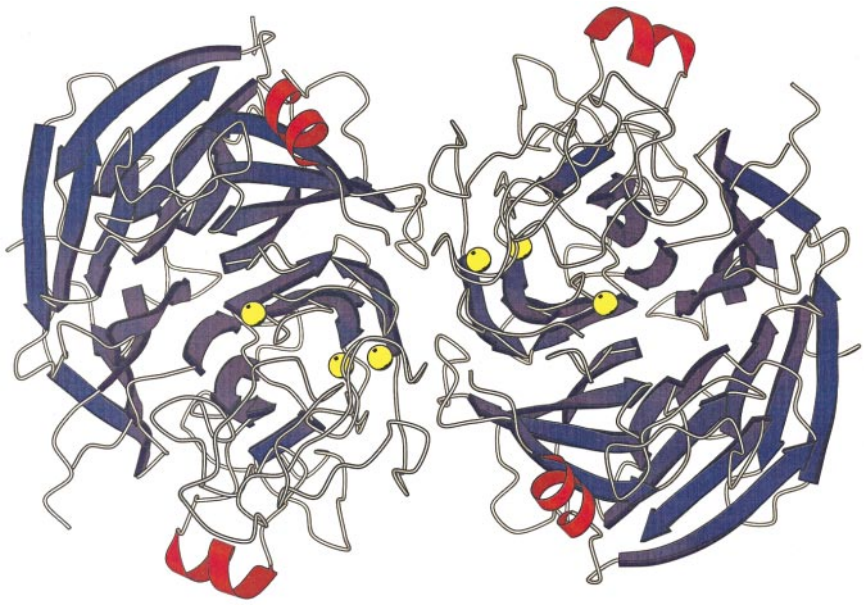

**Figure 2.** Ribbon diagram of the overall structure of the dimer of s-GDH from *Acinetobacter calcoaceticus*. This and all other model pictures were produced with MOLSCRIPT (Kraulis, 1991). β-Strands are shown in blue and α-helices in red. The calcium ions are shown as large, yellow spheres. Coloring is maintained in other Figures unless stated otherwise.

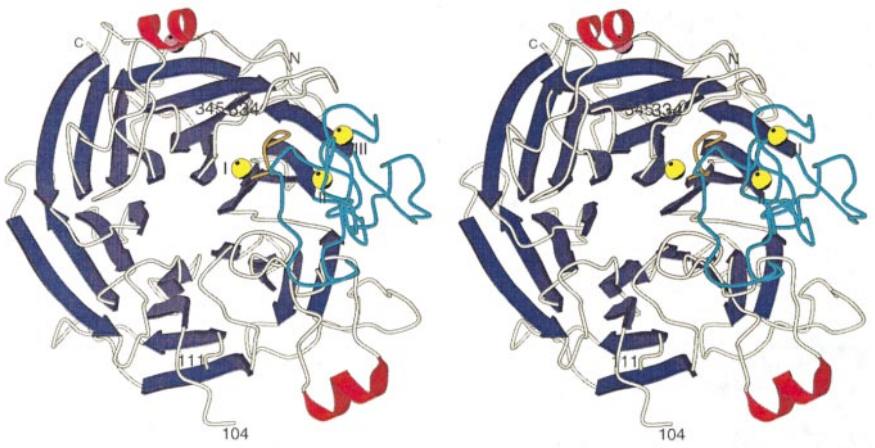


**Figure 3.** Ribbon diagram of one s-GDH monomer. The platinum ion bound to this monomer is shown as a purple sphere. The 4BC and 4CD loops are shown in orange and light blue coils, respectively. The N and C termini, and the residues preceding and following a missing loop are indicated.

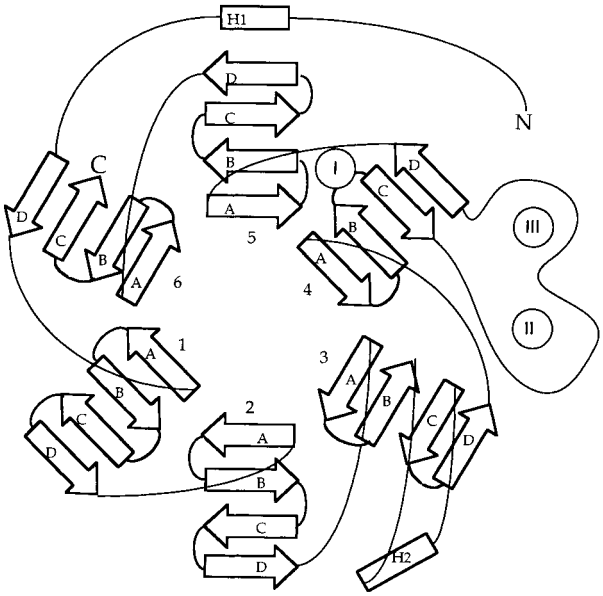


**Figure 4.** Topology diagram showing the arrangement of the secondary structure elements of s-GDH. β-Strands are shown as arrows, α-helices as rectangulars, and the calcium ions as circles. This type of arrangement of β-sheets gives rise to a β-propeller fold.

around a pseudo 6-fold symmetry axis that runs through the center of the molecule, forming a large propeller-like structure. Hence, the name β-propeller has been introduced for this type of protein fold (Murzin, 1992; Varghese *et al.*, 1983). The central channel of the β-propeller, close to the axis of pseudo-symmetry, is filled with numerous water molecules, whereas it is blocked by several amino acid side-chains in the center of the structure.

Following the common terminology of β-propellers (Faber *et al.*, 1995; Ghosh *et al.*, 1995; Renault *et al.*, 1998; Wall *et al.*, 1995; Xia *et al.*, 1992), the β-strands are labelled according to the β-sheet in which they occur (1-6). Within one four-stranded anti-parallel β-sheet, the strands are labelled from the inside to the outside of the molecule (A-D). The sequence strictly follows the down-up-down-up or ''W'' structural topology of the β-sheets from the inside to the outside of the molecule and then *via* a loop to the next β-sheet (Figure 4). The sole exception to this rule is β-sheet 6, whose A, B and C β-strands emerge into the C terminus, whereas the D β-strand is provided by the N-terminal region. The same type of ''Velcro'' closure (Neer & Smith, 1996) has been found in PQQ-dependent methanol dehydrogenase and some other β-propeller proteins, but different types of Velcro closure have been observed in the other β-propellers (Neer & Smith, 1996).

Crystal structures of several enzymes with a β-propeller fold have been reported, which use four to eight β-sheets to construct the barrel. Haemopexin and homologous proteins use four β-sheets to form a β-propeller (Faber *et al.*, 1995; Gohlke *et al.*, 1996; Li *et al.*, 1995; Libson *et al.*, 1995), whereas, like s-GDH, neuraminidases have six blades to construct the barrel (Crennell *et al.*, 1994; Gaskell *et al.*, 1995; Varghese & Colman, 1991; Varghese *et al.*, 1983). A 7-fold pseudo-symmetry is utilized in the β-subunit of G-proteins (Sondek *et al.*, 1996; Wall *et al.*, 1995), copper-containing galactose oxidase (Ito *et al.*, 1991, 1994) and quinoprotein methylamine dehydrogenase (Chen *et al.*, 1992; Vellieux *et al.*, 1989). MDH and nitrite reductase, finally, have β-propellers that consist of eight β-sheets (Fülöp *et al.*, 1995; Ghosh *et al.*, 1995; Xia *et al.*, 1992, 1996).

**Table 3.** Distances between the six Ca ions and their ligands in s-GDH

| $Ca^{2+}$ (*B*-factor (Å$^2$)) | Ligand | Distance to $Ca^{2+}$ (Å) | *B*-factor (Å$^2$) |
|---|---|---|---|
| $Ca^{2+}$ I in monomer A/B (18.3/17.0) | Gly247 O | 2.37/2.35 | 9.1/8.6 |
| | Pro248 O | 2.37/2.34 | 7.3/7.2 |
| | Water425/265 O | 2.40/2.33 | 17.2/12.6 |
| | Water108/140 O | 2.41/2.45 | 9.5/10.3 |
| | Water406/220 O | 2.30/2.41 | 14.0/13.6 |
| | Water807 O | 2.54 | 25.01 |
| $Ca^{2+}$ II in monomer A/B (7.0/7.8) | Glu253 $O^{\varepsilon 1}$ | 2.50/2.51 | 10.1/7.4 |
| | Glu253 $O^{\varepsilon 2}$ | 2.42/2.42 | 8.7/9.3 |
| | Tyr263 O | 2.33/2.24 | 9.5/6.0 |
| | Water109/12 O | 2.32/2.34 | 7.6/6.1 |
| | Water52/32 O | 2.44/2.42 | 9.1/6.2 |
| | Water79/18 O | 2.34/2.34 | 7.0/6.4 |
| | Water70/55 O | 2.45/2.36 | 9.7/4.6 |
| $Ca^{2+}$ III in monomer A/B (8.2/7.8) | Ala 269 O | 2.33/2.26 | 7.0/5.3 |
| | Tyr271 O | 2.31/2.34 | 8.0/8.4 |
| | Asp273 $O^{\delta 1}$ | 2.33/2.34 | 6.2/4.8 |
| | Glu309 $O^{\varepsilon 1}$ | 2.52/2.48 | 11.7/9.2 |
| | Glu309 $O^{\varepsilon 2}$ | 2.51/2.47 | 7.9/6.5 |
| | Water817/29 | 2.48/2.48 | 7.5/5.8 |
| | Water50/1045 | 2.35/2.32 | 4.8/6.2 |

### $Ca^{2+}$-binding sites

s-GDH needs $Ca^{2+}$ for dimerization and activity. In the crystal structure of the apo form of s-GDH, three $Ca^{2+}$-binding sites are present per protein subunit. All three calcium ions have *B*-factors comparable to the immediate protein environment, and are octahedrally coordinated as expected, with five to seven liganding atoms at distances between 2.2 and 2.6 Å (Table 3, Figure 5). One calcium ion (I) is located near the 6-fold pseudo-symmetry axis, at one end of the β-propeller (Figure 3). It is liganded by the two main-chain carbonyl oxygen atoms of Gly247 and Pro248 (located in loop 4BC) and four water molecules (Figure 5(a)). Its position resembles the PQQ-bound $Ca^{2+}$ in MDH (see the section A structure comparison of s-GDH and MDH suggests the active-site location of s-GDH), and therefore it could function in the binding and functionalization of PQQ. The other two calcium ions (II and III) are located in the 4CD loop (Figures 3, and 5(b) and (c)), keeping together sequentially distant parts of the this loop (which comprises residues 258-319): ligands of calcium ion II and/or its liganding water molecules are donated by residues 226, 245, 253, 255, 268, 270 and 318 (Figure 5(b)), while ligands for calcium ion III and/or its liganding water molecules are provided by residues 251, 269, 271, 273, 275 and 309 (Figure 5(c)). The requirement of $Ca^{2+}$ for dimerization may thus be explained by the necessity to rigidify the loop that constitutes the majority of the dimerization interface (see below).

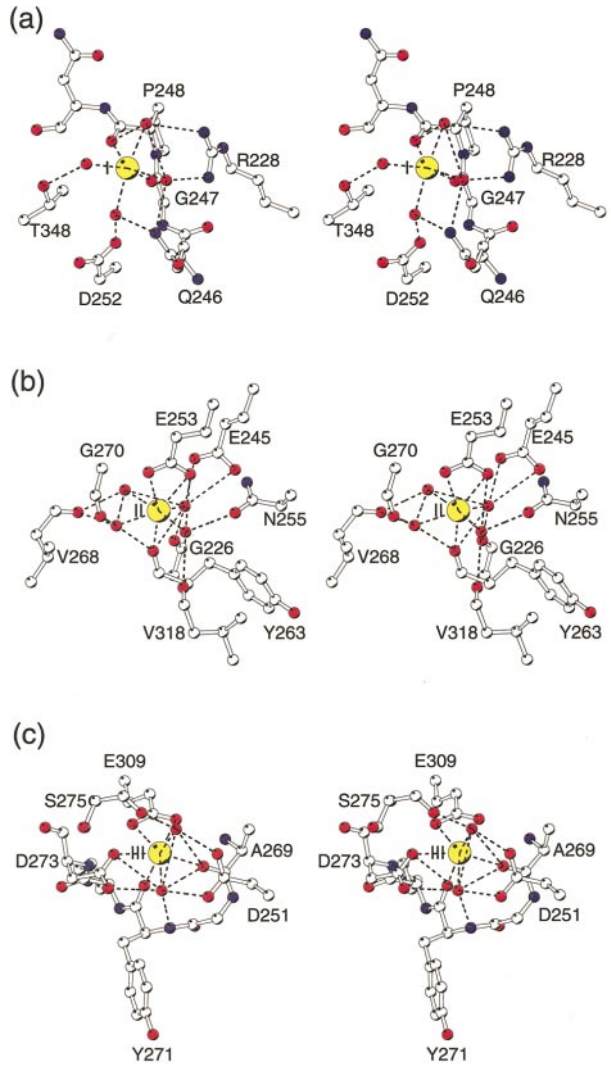


**Figure 5.** Stereo view of the $Ca^{2+}$-sites of s-GDH in ball and stick representation: (a) $Ca^{2+}$-binding in the putative active site; (b) and (c) in the dimerization interface. Carbon, nitrogen, oxygen, and calcium atoms are shown in white, blue, red, and yellow, respectively. Hydrogen-bonding interactions are marked with a broken line.

### Subunit interactions

Holo s-GDH is a dimeric molecule, consisting of two identical subunits of 50 kDa each, as was indicated by a number of different experiments including non-denaturing gel electrophoresis, gel-filtration chromatography and sedimentation-equilibrium centrifugation (Dokter *et al.*, 1986; Geiger & Görisch, 1986; Olsthoorn & Duine, 1996; Olsthoorn *et al.*, 1997). The dimer is destabilized with increasing temperature and/or pH, whereas the addition of NaCl has a stabilizing effect (Olsthoorn *et al.*, 1997). Because of the counteracting forces of high pH and the presence of NaCl, it cannot be predicted *a priori* whether the dimer is stable or not under the conditions of the crystallization experiments (3 mM $CaCl_2$, 120 mM NaCl, pH 9.2).

We performed dynamic light-scattering experiments to determine the oligomerization state of s-GDH under crystallization conditions. Polyethyleneglycol (PEG) 6000 increased the polydispersity and residual error of the samples and was therefore removed from the protein solutions. A molecular mass of 71 kDa was obtained from these experiments when a calibration model for globular proteins was used (see Materials and Methods). This value is similar to that obtained from gel-filtration chromatography (Olsthoorn & Duine, 1996). However, a molecular mass of 92 kDa (close to 100 kDa for a dimer) was obtained when a model for non-globular proteins was used.

Under crystallization conditions, PEG6000 is present as a precipitant, but no such molecules were found in the dimer interface, and therefore PEG6000 is not expected to directly interfere with dimer formation. However, during equilibration of the drop and the reservoir, water evaporates from the drop to the reservoir, thus increasing the protein concentration in the drop. Since the equilibrium constant $[monomer]^2/[dimer]$ remains the same, an increase of the protein concentration results in a shift of the monomer/dimer equilibrium towards formation of the dimer. It is thus likely that s-GDH is dimeric in the crystallization set-ups and that the functional dimer is present in the crystal structure.

**Table 4.** Hydrogen-bonding interactions in the dimerization interface

| Distances (Å) (AB)/(BA) | Atoms in monomer A/B | Atoms in monomer B/A |
|---|---|---|
| 2.9/2.9 | Lys272 NZ | Asp396 OD1 |
| 3.4/3.4 | Lys272 NZ | Asp396 OD2 |
| 3.1/3.1 | Asp274 O | Gln328 NE2 |
| 3.3/3.4 | Gly314 O | Asp395 OD1 |
| 3.1/3.1 | Asn316 N | Asp395 OD1 |
| 2.8/2.9 | Asn316 N | Asp395 OD2 |
| 2.9/2.8 | Phe317 N | Asp395 OD1 |
| 3.4/3.3 | Phe317 O | Thr393 O |
| 2.9/3.0 | Phe317 O | Asp395 N |
| 2.8/2.8 | Gln328 OE1 | Asp329 N |
| 3.3/3.2 | Gln328 NE2 | Asp329 OD2 |

The largely planar dimer interface covers an area of roughly 900 $A^2$ and comprises residues 1-2 (N terminus), 254-259 (β-strand 4C, loop 4CD), 271-275 (loop 4CD), 310-330 (loop 4CD, β-strand 4D, loop 4D5A) and 387-396 (loop 5CD) of both protein monomers. There are numerous direct protein-protein interactions, most of which are hydrophilic and more than half of the interactions involve main-chain atoms. Hydrogen-bonding interactions involve the side-chains of Lys272, Gln328, Asp329, Asp395 and Asp396 (Table 4). Numerous water molecules are present in the dimerization interface, mediating many hydrogen-bonds. The dimerization interface is thus stabilized mainly by hydrophilic interactions, which is in good agreement with the influence of NaCl and temperature on the stability of the dimer. Moreover, since the dimer interface is covered with mostly polar and negatively charged residues (Figure 6), the pH-dependence of dimer stability can be explained by the accumulation of negative charge in the interface with increasing pH. Above pH 8 in the absence of salt, the monomers repel each other, which results in destabilization of the dimer and the formation of monomers.

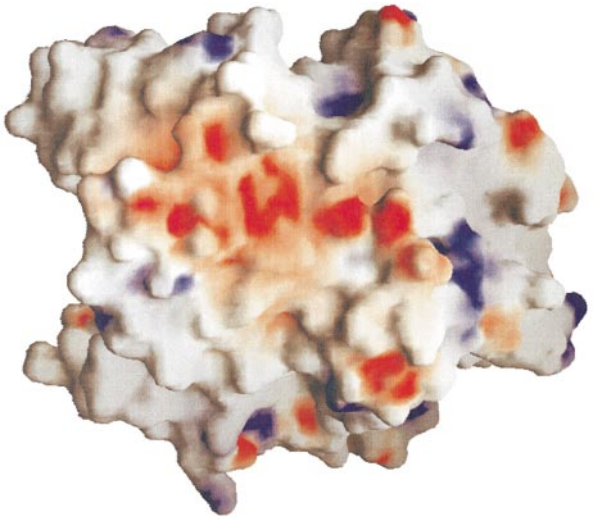

**Figure 6.** Molecular surface drawing of one s-GDH monomer created using the program GRASP (Nicholls & Honig, 1993), showing the large negatively charged patch in the dimerization interface. Neutral, positively and negatively charged patches are shown in gray, blue, and red, respectively.

## A structure comparison of s-GDH and MDH suggests the active-site location of s-GDH

s-GDH is a β-propeller protein with six β-sheets (Figures 3 and 7(a)). The only other PQQ-dependent enzymes with known three-dimensional structures are two bacterial methanol dehydrogenases (MDHs), which have an $\alpha_2\beta_2$ subunit composition (Figure 7(b)). The α-domains adopt a β-propeller fold consisting of eight four-stranded anti-parallel β-sheets. PQQ and $Ca^{2+}$ are bound near the top of this structural motif, close to the pseudo 8-fold symmetry axis formed by the eight β-sheets. Only one stretch of 31 amino acid residues (167-197) in s-GDH was reported to have significant sequence identity with other PQQ-dependent proteins (Cleton-Jansen *et al.*, 1990). In s-GDH, this part of the sequence forms an irregular loop with one α-helix at the surface of the protein (Figure 7(a)), whereas it forms β-strands 7D, 8A and 8B as well as the loop between β-sheets 7 and 8 in MDH. It is thus involved in the formation of the β-propeller in MDH (Figure 7(b)). Thus, although there is homology in a short stretch of the sequences of s-GDH and other PQQ-containing proteins, these stretches adopt distinct conformations. On the basis of a structure comparison of s-GDH and MDH, the observed sequence identity in a stretch of 31 residues seems fortuitous.

By analogy of the active site location of MDH and other β-propeller structures, the PQQ-binding site of s-GDH is expected to reside at the top of the barrel near the pseudo-symmetry axis. However, no electron density was observed to account for the presence of PQQ in this crystal structure. Activity measurements showed that dissolved crystals were inactive but that activity was restored instantly upon addition of PQQ, confirming the absence of PQQ in the crystals. It appeared that Tris, which was used as a buffer in the reconstitution and crystallization solutions, can react with PQQ (van Kleef & Duine, 1989). When incubated in Tris-buffer for more than one day, s-GDH loses activity (Olsthoorn & Duine, 1996). Nevertheless, our crystal structure presents some clues to corroborate the proposed location of the PQQ-binding site. Near the top of the barrel a deep, wide and positively charged cleft is present in s-GDH that seems large enough to accommodate PQQ. In this cleft, one calcium ion is available that could function in the binding and functionalization of PQQ, in agreement with biochemical evidence. Moreover, in MDH a calcium ion is present in a similar position, close to PQQ, as well. Hence, it seems likely that the active site of s-GDH is located at the top of the barrel near the pseudo-symmetry axis. Recently, we succeeded in binding PQQ in an

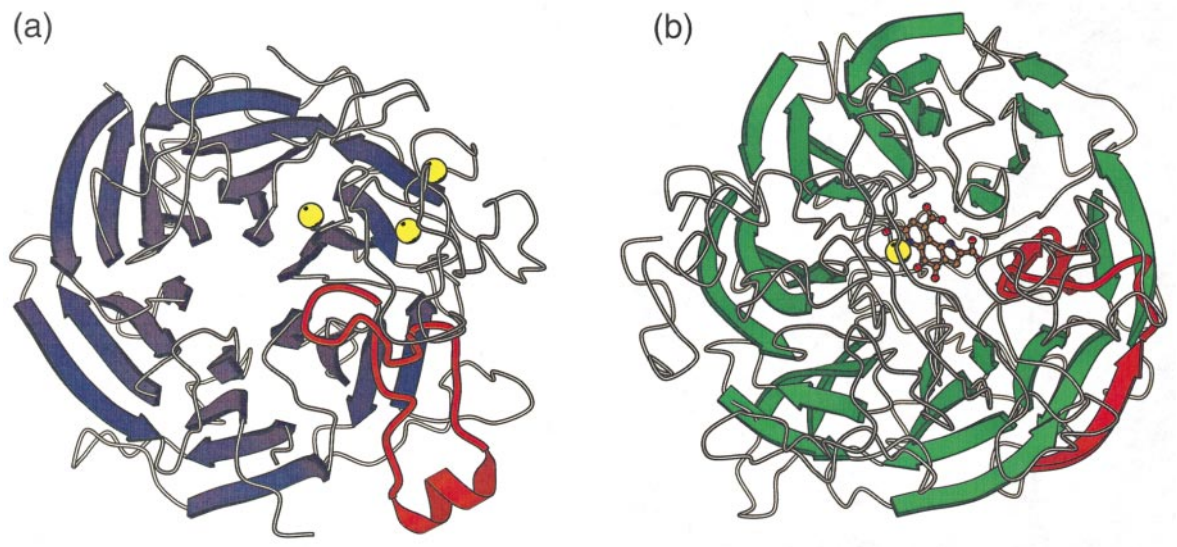


**Figure 7.** Structure comparison of (a) one s-GDH monomer and (b) one α-subunit of MDH. β-Strands of s-GDH and MDH are shown as blue and green arrows, respectively. The remainder of the structures is shown as a gray coil. PQQ is shown in ball and stick representation. Homologous parts in the sequences of both proteins are shown in red. The structures were manually aligned with the axes of pseudo-symmetry perpendicular to the plane of the picture, and with the conserved sequences in a similar position.

s-GDH crystal. The position of PQQ confirms the location of the active site proposed here (unpublished results).

## Comparison with homologous sequences

Although previous data base searches had not revealed any proteins homologous to s-GDH, a new search with the program BLAST (Altschul *et al.*, 1997) surprisingly showed that two open reading frames from the complete genomes of *Escherichia coli* K-12 strain MG1655 (371 amino acid residues, 41 kDa: Blattner *et al.*, 1997) and *Synechocystis sp.* strain PCC6803 (412 amino acid residues, 45 kDa: Kaneko *et al.*, 1996) as well as two incomplete sequences from the genomes of *Pseudomonas aeruginosa* (343 amino acid residues) and *Bordetella pertussis* (247 amino acid residues) are homologous to s-GDH. Several protein localization programs such as PSORT (Nakai & Kanehisa, 1991) and SignalP (Nielsen *et al.*, 1997) indicate that the *E. coli* enzyme may be located in the periplasm, whereas the *Synechocystis sp.* enzyme may be either periplasmic or attached to the cytoplasmic membrane *via* a transmembrane helix/linker. The N-terminal part of the *Pseudomonas* and *Bordetella* sequences were too short to assess their localization.

In the following analysis, the *Bordetella* sequence was not included, because it encodes only part of the structural core of s-GDH. A structure-guided multiple sequence alignment of the four remaining protein sequences indicates that 19 % of the residues are identical (Figure 8). The N-terminal parts of the sequences, which (may) include a signal peptide, have different lengths and show only weak homology. Significant sequence similarity starts at residue Leu26 at the beginning of the 1A strand in s-GDH, and extends up to residue 448, at the end of the 6C strand. A total of ten hydrophobic core residues and 19 glycine and proline residues, located at turning points in the structure, are absolutely conserved. Moreover, several residues that give rise to special structural features in s-GDH, such as the stacking of the side-chains of Trp265 and *cis*-Pro266, the π/π interaction between Tyr357 and Trp367, and many residues of the structural motifs of the internal sequence repeat (see the next paragraph) are also absolutely conserved. Strikingly, a significant part of the dimerization interface, including a substantial part of the long 4CD loop, has been deleted. Moreover, the calcium binding residues in loop 4CD, which are important for dimerization, are not conserved. This indicates either that the three recently discovered potential proteins are monomeric or that their dimerization interfaces may be $Ca^{2+}$-independent and totally different from the interface of s-GDH. As the structural core of the β-propeller seems to be conserved in the four homologous protein sequences, and gaps in the alignment can be explained by the insertion or truncation of loop regions, we suggest that these proteins all have a β-propeller fold consisting of six β-sheets.

Apart from the residues conserved for structural reasons, a large number of residues is absolutely conserved in the putative active site. These residues include Gln76, His144, Arg228, Asn229, Gln231, Gly247, the carbonyl oxygen atom of which is a ligand of the ''active-site calcium ion'', Asp252, Ala350, Leu376, Arg406, Arg408, and Asp424. The high degree of side-chain conservation in this protein cleft suggests that the three proteins may bind PQQ as well and that they may catalyze a similar reaction.

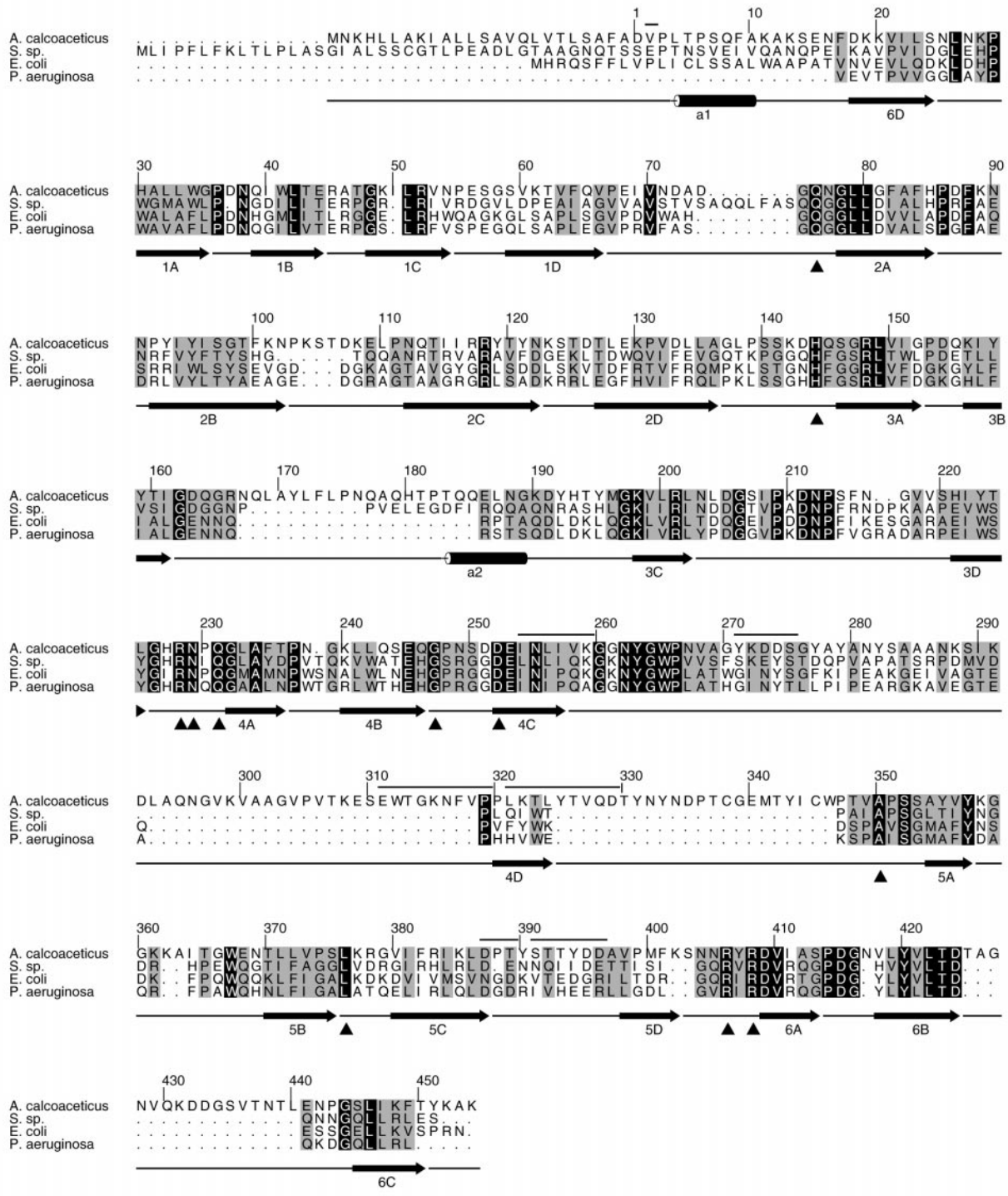


**Figure 8.** Structure-based sequence alignment of GDH and three homologous sequences from *E. coli, P. aeruginosa* and *Synechocystis sp*. Identical and similar residues have black and gray boxes, respectively. Conservation values, ranging from 0 to 10 were calculated for all amino acid positions, with 10 being identity (Barton, 1993). Positions with values above 5 show increasing similarities in physico-chemical properties and were designated as similar. Absolutely conserved residues in the presumed active site are marked with a triangle. Residues present in the dimer interface are indicated by a line above the s-GDH sequence.

## The internal sequence repeat

In s-GDH, six topologically identical β-sheets are present. β-Sheets 2-6 were superimposed on β-sheet 1 with rms fits between corresponding $C^{\alpha}$ atoms ranging from 1.3-1.6 Å for 30-37 atoms (Figure 9(a)). In general, β-strands A, B and C superimpose well, whereas more variation is observed in the solvent-exposed D β-strands. The β-sheets 4 and 5 deviate most, presumably because

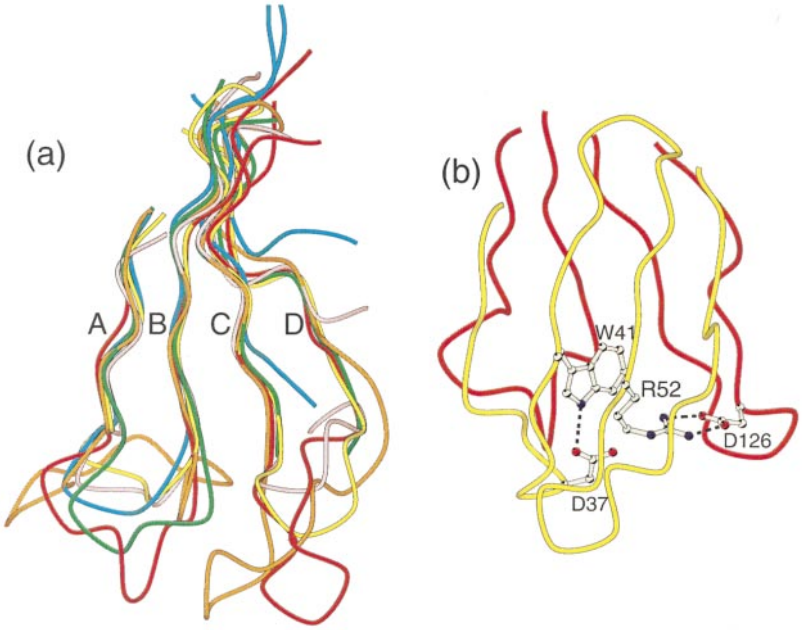


**Figure 9.** (a) Superposition of the six β-sheets of the three-dimensional structure of s-GDH. The six β-sheets are shown in different colors. Some loops connecting two β-strands have been removed for clarity. (b) The stabilizing interactions of the structural motifs between β-sheets 1 (yellow) and 2 (red) in s-GDH. Hydrogen-bonding interactions are marked with a dotted line.

they are involved in the binding of the three calcium ions and are located close to the dimerization interface.

The superposition of the six β-sheets allows a structure-based alignment of the amino acid sequences (Figure 10). Although the degree of sequence identity between the six β-sheets is rather low, there is a clear preference for hydrophobic (Val, Phe, Ala, Gly, Ile and Leu) side-chains at β-strand positions A3-5, B2, B4, C2-3, D2-4, and D6. Moreover, the consensus sequence L-x(3)-PDG-x(0,10)-I/L-Y-x(6)-G-x(1,33)-I-x-R-x(3,7)-D can be recognized (a residue is judged conserved if it is present at least three times at a certain position in the sequence alignment). The hydrophobic residues in the consensus sequence form the core of the structure together with the other residues at positions where hydrophobic side-chains are preferred. The conserved Pro and Gly are located in the AB loop to allow kinks in the peptide backbone. The conserved Gly in the BC loop has a similar function. In addition, the hydrophilic residues in the consensus sequence give rise to two distinct structural motifs. Motif I involves the Asp at position 2 in the AB loop and the Tyr at position 3 in β-strand B. The second structural motif comprises the Arg/Lys at position 4 in β-strand C and the Asp typically four residues ahead in the CD loops.

Motif I is found in the short β-turn between β-strands A and B. The Asp points back into the β-sheet to form a hydrogen-bond with the OH/NH of the aromatic side-chain at position three in β-strand B (Figure 9(b)). This hydrogen-bond is constructed only in β-sheets 1 and 3. The details of the hydrogen-bonding pattern in the other β-sheets are slightly different. In β-sheet 2, a possible hydrogen-bond between Asp87 and Tyr95 is mediated by the side-chain of His85. In β-sheet 4, Asn238 occupies the position of the typical Asp, forming an alternative hydrogen-bond with Thr236 of the A β-strand in the same β-sheet. In β-sheet 5, this motif is absent, and Tyr415 and Asp420 in β-sheet 6 are too far apart for hydrogen-bond formation.

Motif II comprises an Arg/Lys in the C β-strand of one β-sheet, which forms a (bidentate) salt-bridge with the Asp/Glu in the CD-loops of the next β-sheet (Figure 9(b)). This salt-bridge is possible between β-sheets 1 and 2, 2 and 3, and 6 and 1, and a similar salt-bridge is formed between Arg201 and Asp211, which is located in CD-loop of the same β-sheet 3. Arg383 (β-sheet 5) is incapable of making a favorable interaction with the CD loop in the next β-sheet, because this CD loop does

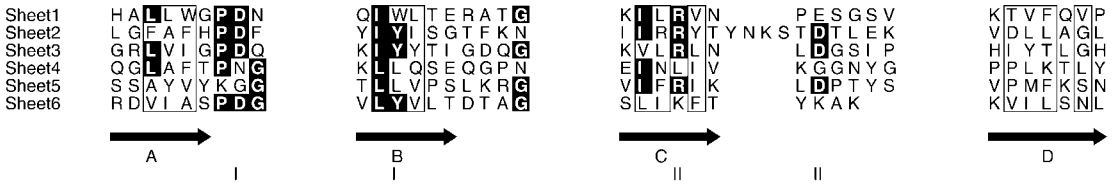


**Figure 10.** Structure-based sequence alignment of the six four-stranded β-sheets in the s-GDH structure. Positions with a preference for hydrophobic residues are boxed and residues with an occurrence of above three are boxed with a black background (produced with ALSCRIPT (Barton, 1993)). The roman symbols I and II indicate the residues forming structural motifs I and II, respectively.

not exist (β-strand C emerges into the C terminus, whereas the D β-strand is provided by the N-terminal region).

At first sight, the internal sequence identity of the β-sheets may seem too low to be significant. On the other hand, closer inspection of the multiple sequence alignment shows that most residues of the consensus sequence have an occurrence of more than 50% in 27 to 30 β-sheets (six β-sheets; four/five protein sequences, including the *B. pertussis* sequence). A similar degree of homology of a few residues per β-sheet has been observed in a number of other β-propellers (discussed in more detail by Baker *et al.*, 1997).

The conserved residues in the internal sequence repeat make hydrophobic and hydrophilic interactions that may provide additional stability to the protein structure. An equivalent function has been suggested for the internal sequence repeats that have been found in some other superbarrels (Faber *et al.*, 1995; Ghosh *et al.*, 1995). The presence of structural motif I in s-GDH may be explained by assuming a gene multiplication event from an ancestral β-sheet containing this motif. Divergent evolution may have resulted in the alteration and/or loss of some conserved residues in some of the β-sheets. A similar evolutionary event has been proposed for other superbarrel proteins (Bork & Doolittle, 1994). From this point of view, the β-sheet multiplication event should have occurred prior to the divergence of the bacteria involved, because the internal sequence identity is smaller than it is among the protein sequences from different bacteria. In contrast, structural motif II extends over two successive β-sheets, and therefore it cannot have been functional in a single ancestral β-sheet. Instead, the presence of this latter motif might be the result of the need to stabilize the β-propeller structure. A similar explanation for the presence of internal repeats in β-propellers has been put forward by Murzin (1992), who argued that the structure of such proteins is dictated by general features of β-sheet structure such as periodicity, twist and amino acid composition. A comparison of the structures of MDH and cytochrome *cd*1 nitrite reductase, which both adopt a β-propeller structure with eight β-sheets, indicates these proteins do not contain the same repeats (Baker *et al.*, 1997). Therefore, the authors concluded from the structure comparison that the presence of the internal repeats is not essential for the fold of the proteins and that it seems unlikely that all β-propeller proteins are derived from one common ancestral β-sheet comprising all known internal sequence repeats. However, this does not exclude the possibility that the β-propellers have been formed from different ancestral β-sheets, giving rise to the various internal sequence repeats.

The two structural motifs that emerge from the crystal structure of s-GDH are not observed in any of the other superbarrel proteins, although motif I shows some structural resemblance to the hydrogen-bonding network formed in G-proteins, with an Asp in a loop pointing back into the β-sheet to form a hydrogen-bond. The hydrogen-bonding network in the latter class of proteins, however, is much more elaborate and the location in the sequence of the amino acids participating in hydrogen-bond formation is completely different. Furthermore, the Asp in β-strand 3, which is conserved in several PQQ-dependent dehydrogenases, G-proteins, neuraminidases, and cytochrome *cd*1 nitrite reductase, is not present in s-GDH. Thus, s-GDH presents two novel types of interactions to stabilize its architecture, thereby adding novel information for the discussion of the evolution of β-propeller proteins.

## Conclusions

s-GDH adopts a β-propeller fold with six antiparallel four-stranded β-sheets aligned around a 6-fold pseudo-symmetry axis. The dimeric enzyme binds three calcium ions per monomer, two of which are present in the dimerization interface, and a third is located in the putative active site. Unfortunately, the PQQ cofactor is absent in the crystal structure presented here, due to the presence of Tris in the crystallization medium. Structural details on the catalytic mechanism of quinoproteins is scarce. Further research is required to determine cofactor and substrate binding and to analyze the different redox states of the enzyme.

An internal structure-based sequence alignment shows the presence of a conserved sequence repeat that gives rise to two structural motifs with a stabilizing function: the first motif is found at the corner of the short β-turn between β-strands A and B, with an Asp side-chain in the AB loop forming a hydrogen-bond with the OH/NH of a side-chain in the B β-strand of the same β-sheet. The second motif comprises an Arg/Lys in the C β-strand of β-sheet $n$, which forms a bidentate salt-bridge with an Asp/Glu in the CD-loop of β-sheet $n + 1$.

s-GDH is not a unique enzyme, as has long been thought. At least three potential protein sequences are homologous to s-GDH and are consequently expected to have a similar β-propeller fold. A structure-based sequence alignment showed that several side-chains are absolutely conserved in the putative active site, indicating that these proteins may all be able to bind and functionalize PQQ and possibly catalyze similar reactions. Characterization of the potential protein sequences in terms of localization, substrate specificity and reactivity towards natural electron acceptors may provide insight into the unknown function of s-GDH and its homologs *in vivo*.

## Materials and Methods

### Purification and crystallization

Recombinant s-GDH was kindly provided in a purified, lyophilized state by Boehringer Mannheim

(Germany). Since *E. coli* does not produce free PQQ, the enzyme is produced in the apo form (Hommes *et al.*, 1984; Olsthoorn & Duine, 1996). Therefore, the enzyme was first solubilized and reconstituted in a solution containing 0.3 mM PQQ and 3 mM $Ca^{2+}$. Because degradation of the lyophilized protein material occurs due to aging, the protein material was purified directly prior to crystallization on a MONO S HR5/5 cation-exchange column. Reconstituted s-GDH was applied to the column in buffer A (20 mM Tris/glycine (pH 8.5), 3 mM $CaCl_2$) and eluted using a linear gradient from 0 M to 1 M NaCl in buffer A. The temperature was maintained at 283 K during the entire purification procedure. The non-bound fractions and fractions from 0.10 to 0.12 M NaCl were pooled and concentrated to 10 mg/ml by ultra-filtration using a Centricon-30 concentrator (Amicon). Crystallization experiments were based on conditions described by Schlunegger *et al.* (1993). The protein solution was mixed with 0.1 mM PQQ and 8% (w/v) PEG6000 in 50 mM Tris/glycine (pH 9.2), obtaining a protein concentration of 5 mg/ml. This solution was subsequently centrifuged for ten minutes: 8 µl of the centrifuged solution was equilibrated against 20-23% (w/v) PEG6000, 120 mM NaCl, 3 mM $CaCl_2$, 50 mM Tris/glycine (pH 9.2). As a result of the purification step prior to crystallization, the crystal size and quality improved dramatically compared to crystals that were obtained with the method according to Schlunegger *et al.* (1993). The largest and best-diffracting crystals were obtained by adding a seed crystal in a solution containing 5 mg s-GDH/ml, 13% (w/v) PEG6000, 3 mM $CaCl_2$, 120 mM NaCl, 50 mM Tris/glycine (pH 9.2). This solution was first equilibrated against 16% (w/v) PEG6000, 3 mM $CaCl_2$, 120 mM NaCl, 50 mM Tris/glycine (pH 9.2).

s-GDH crystals are very stable in an artificial mother liquor consisting of 23% (w/v) PEG6000, 120 mM NaCl, 3 mM $CaCl_2$, 50 mM Tris/glycine (pH 9.2). The dramatic improvement in crystal stability can be attributed to the higher purity of the protein sample combined with an ionic strength in the reservoir and stabilizing solutions that optimally approximates the conditions of the drop. The resulting crystals belonged to space group $P2_1$ with cell dimensions $a = 61.9$ Å, $b = 94.2$ Å, $c = 86.5$ Å, $\beta = 104.9°$, and with two monomers in the asymmetric unit. A cryoprotectant solution containing 15% (v/v) glycerol, 25% (w/v) PEG6000, 120 mM NaCl, 3 mM $CaCl_2$, 50 mM Tris/glycine (pH 9.2), can be used to freeze the crystals. However, s-GDH crystals are not stable in this solution, requiring the soaking time of the crystals in the cryo-solution to be limited as much as possible. Typical soaking times were shorter than half a minute. Compared to room temperature, frozen crystals have slightly reduced cell dimensions: $a = 60.2$ Å, $b = 92.1$ Å, $c = 84.9$ Å, $\beta = 105.2°$.

## Data collection

A native data set and three derivative data sets were collected at room temperature on an in-house MacScience DIP2000 image plate (Enraf Nonius, Delft, The Netherlands) connected to an Elliot GX 21 rotating anode generator. One derivative data set was collected at room temperature on the MAR Research image plate system at the X-31 beam line of the EMBL Outstation at DESY, Hamburg, and a 1.72 Å resolution data set of a $Pt^{2+}$-derivatized crystal was collected under cryogenic conditions at the X-11 beam line of the EMBL Outstation at DESY, Hamburg. All data were processed and reduced with the DENZO/SCALEPACK package (Otwinowski & Minor, 1997) and software of the Groningen BIOMOL package was used to scale all data sets together (Table 1).

## Structure determination

All difference Patterson and difference Fourier maps were calculated with programs from the PHASES package (Furey & Swaminathan, 1997). Heavy-atom parameter and phase refinement were done with the same package (Table 1). During refinement, contributions to the protein phase probability were excluded from the data set for which parameters were being refined to minimize protein phase bias from the heavy-atom parameters (Blow & Matthews, 1973; Furey & Swaminathan, 1997). The latter procedure resulted in a slight breakdown of the phasing statistics as compared to using all data during heavy-atom parameter refinement, but the quality of electron density maps improved significantly. A solvent-flattening procedure of the PHASES package, consisting of several alternating rounds of boundary determination, solvent flattening and phase combination, was used to improve the initial phases. At this point, a 3.0 Å electron density map allowed the unambiguous tracing of approximately 300 out of 454 $C^{\alpha}$ positions per monomer.

The NCS matrix, relating the two molecules in the asymmetric unit, was found by simply examining symmetry in the heavy-atom positions in the crystallographic cell. A mask was calculated on the basis of the polyalanine model that was available from the $C^{\alpha}$ trace. As a result of the NCS averaging with the program DM (Cowtan & Main, 1993), the quality of electron density maps improved considerably. As the tracing proceeded and the polyalanine model was extended, several new masks were calculated on the basis of the latest model and used in the calculation of new phases, thus improving the protein phases in an iterative manner.

## Refinement

The model was initially refined using a simulated annealing protocol from X-PLOR, version 3.843 (Brünger *et al.*, 1987) using data to 2.7 Å resolution collected at room temperature. A test set of 5% of the unique reflections was used to calculate a free *R*-factor (Brünger, 1992). This was followed by rigid body refinement at 5.0 Å resolution using the cryo-data in order to find the proper orientation of the dimer in the cell under these conditions. One round of refinement using simulated annealing at 2.7 Å was done to overcome possible conformational differences in the structures at room and cryo-temperature.

At this point, the strong NCS restraints that had been imposed on the model in previous refinement cycles were released and several rounds of an automated refinement procedure were carried out at 1.72 Å, alternated with manual rebuilding steps. The automated refinement procedure was performed with REFMAC, ARP and auxiliary programs from the CCP4 package (Dodson *et al.*, 1997), and consisted of performing iterative cycles of maximum likelihood positional and individual *B*-factor refinement, followed by removal of atoms that were out of density, and subsequent addition of water molecules in those parts of electron density, that were not accounted for by the current model. In the final stage of refinement, the bulk solvent correction and anisotropic scaling procedures of X-PLOR were used to

optimise the model further. Refinement statistics are summarized in Table 2.

### Dynamic light-scattering experiments

Dynamic light-scattering experiments were performed using a Dyna Pro-801 instrument (Protein Solutions, Charlottesville, VA). The protein solution consisted of 5 mg/ml s-GDH, 3 mM $CaCl_2$, 120 mM NaCl, and 36/39 mM Tris/glycine (pH 9.2). The measured hydrodynamic radius was 3.74 nm at 393 K. Different molecular masses were calculated depending on what molecular mass model was used. With the standard size/mass calibration model for globular proteins, a molecular mass of 71 kDa was calculated, whereas a molecular mass of 92 kDa was obtained with the volume shape hydration/molecular mass model, assuming a partial specific volume of 0.726 ml/g and a frictional ratio of 1.26. These latter two parameter values are averages of several non-globular proteins with molecular masses between 24 and 110 kDa.

### Protein Data Bank accesion numbers

The structure factors and coordinates of the structure of the apo-form of *A. calcoaceticus* s-GDH have been deposited with the Rutgers Protein Data Bank (accession code 1QBI).

## Acknowledgments

We gratefully thank the staff of the EMBL outstation at DESY, Hamburg, Germany, for their help during data collection, and the European Union for support of the work at the EMBL Hamburg through the HCMP Access to Large Installations Project, contract number CHGE-CT93-0040. The investigations were supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO).

## References

Altschul, S. F., Madden, T. L., Schiffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein data base search programs. *Nucl. Acids Res.* **25**, 3389-3402.

Baker, S. C., Saunders, N. F. W., Willis, A. C., Ferguson, S. J., Hajdu, J. & Fülöp, V. (1997). Cytochrome *cd*1 structure: unusual haem environments in a nitrite reductase and analysis of factors contributing to β-propeller folds. *J. Mol. Biol.* **269**, 440-455.

Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37-40.

Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science,* **277**, 1453-1474.

Blow, D. M. & Matthews, B. M. (1973). Parameter refinement in the multiple isomorphous-replacement method. *Acta Crystallog. sect. A,* **29**, 56-62.

Bork, P. & Doolittle, R. F. (1994). *Drosophila kelch* motif is derived from a common enzyme fold. *J. Mol. Biol.* **236**, 1277-1282.

Brünger, A. T. (1992). Free *R* value: a novel statistical quantity for assessing the accuracy of structures. *Nature,* **355**, 472-475.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic *R*-factor refinement by molecular dynamics. *Science,* **235**, 458-460.

Chen, L., Mathews, F. S., Davidson, V. L., Huizinga, E. G., Vellieux, F. M. & Hol, W. G. (1992). Three-dimensional structure of the quinoprotein methylamine dehydrogenase from *Paracoccus denitrificans* determined by molecular replacement at 2.8 Å resolution. *Proteins: Struct. Funct. Genet.* **14**, 288-299.

Cleton-Jansen, A.-M., Goosen, N., Wenzel, T. J. & van de Putte, P. (1988). Cloning of the gene encoding quinoprotein glucose dehydrogenase from *Acinetobacter calcoaceticus*: evidence for the presence of a second enzyme. *J. Bacteriol.* **170**, 2121-2125.

Cleton-Jansen, A.-M., Goosen, N., Fayet, O. & van de Putte, P. (1990). Cloning, mapping, and sequencing of the gene encoding *Escherichia coli* quinoprotein glucose dehydrogenase. *J. Bacteriol.* **172**, 6308-6315.

Cowtan, K. D. & Main, P. (1993). Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. *Acta Crystallog. sect. D,* **49**, 148-157.

Crennell, S., Garman, E., Laver, G., Vimr, E. & Taylor, G. (1994). Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure,* **2**, 535-544.

Dodson, E. J., Winn, M. & Ralph, A. (1997). Collaborative computational project, number 4: providing programs for protein crystallography. *Methods Enzymol.* **277**, 620-633.

Dokter, P., Frank, Jzn J. & Duine, J. A. (1986). Purification and characterization of quinoprotein glucose dehydrogenase from *Acinetobacter calcoaceticus* L.M.D. 79.41. *Biochem. J.* **239**, 163-167.

Dokter, P., van Wielink, J. E., van Kleef, M. A. & Duine, J. A. (1988). Cytochrome *b*-562 from *Acinetobacter calcoaceticus* L.M.D. 79.41. Its characteristics and role as electron acceptor for quinoprotein glucose dehydrogenase. *Biochem. J.* **254**, 131-138.

Duine, J. A. (1991). Quinoproteins: enzymes containing the quinonoid cofactor pyrroloquinoline quinone, topaquinone or tryptophan-tryptophan quinone. *Eur. J. Biochem.* **200**, 271-284.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A,* **47**, 392-400.

Faber, H. R., Groom, C. R., Baker, H. M., Morgan, W. T., Smith, A. & Baker, E. N. (1995). 1.8 Å crystal structure of the C-terminal domain of rabbit serum haemopexin. *Structure,* **3**, 551-559.

Fülöp, V., Moir, J. W., Ferguson, S. J. & Hajdu, J. (1995). The anatomy of a bifunctional enzyme: structural basis for reduction of oxygen to water and synthesis of nitric oxide by cytochrome *cd*1. *Cell,* **81**, 369-377.

Furey, W. & Swaminathan, S. (1997). PHASES-95: a program package for processing and analyzing diffraction data from macromolecules. *Methods Enzymol.* **277**, 590-620.

Gaskell, A., Crennell, S. & Taylor, G. (1995). The three domains of a bacterial sialidase: a β-propeller, an

immunoglobulin module and a galactose-binding jelly-roll. *Structure,* **3**, 1197-1205.
Geiger, O. & Görisch, H. (1986). Crystalline quinoprotein glucose dehydrogenase from *Acinetobacter calcoaceticus*. *Biochemistry,* **25**, 6043-6048.
Geiger, O. & Görisch, H. (1989). Reversible thermal inactivation of the quinoprotein glucose dehydrogenase from *Acinetobacter calcoaceticus*. $Ca^{2+}$ ions are necessary for re-activation. *Biochem. J.* **261**, 415-421.
Ghosh, M., Anthony, C., Harlos, K., Goodwin, M. G. & Blake, C. (1995). The refined structure of the quinoprotein methanol dehydrogenase from *Methylobacterium extorquens* at 1.94 Å. *Structure,* **3**, 177-187.
Gohlke, U., Gomis-Rüth, F.-X., Crabbe, T., Murphy, G., Docherty, A. J. & Bode, W. (1996). The C-terminal (haemopexin-like) domain structure of human gelatinase A (MMP2): structural implications for its function. *FEBS Letters,* **378**, 126-130.
Hoenes, J. (1993). Colorimetric assay by enzymatic oxidation in the presence of an aromatic nitroso or oxime compound. In *US Patent 5206147*, Boehringer Mannheim GmbH, United States of America.
Hoenes, J. (1994). Process and agent for the colorimetric determination of an analyte by means of enzymatic oxidation. In *US Patent 5334508*, Boehringer Mannheim GmbH, United States of America.
Hoenes, J. & Unkrig, V. (1996). Method for the calorimetric determination of an analyte with a PQQ-dependent dehydrogenase. In *US Patent 5484708*, Boehringer Mannheim GmbH, United States of America.
Hommes, R. J. W., Postma, P. W., Neijssel, O. M., Tempest, D. W., Dokter, P. & Duine, J. A. (1984). Evidence of a quinoprotein glucose dehydrogenase apoenzyme in several strains of *Escherichia coli*. *FEMS Microb. Letters,* **24**, 329-333.
Ito, N., Phillips, S. E. V., Stevens, C., Ogel, Z. B., McPherson, M. J., Keen, J. N., Yadav, K. D. S. & Knowles, P. F. (1991). Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature,* **350**, 87-90.
Ito, N., Phillips, S. E. V., Yadav, K. D. S. & Knowles, P. F. (1994). Crystal structure of a free radical enzyme, galactose oxidase. *J. Mol. Biol.* **238**, 794-814.
Jabs, A., Weiss, M. S. & Hilgenfeld, R. (1999). Non-proline *cis* peptide bonds in proteins. *J. Mol. Biol.* **286**, 291-304.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A,* **47**, 110-119.
Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A. & Nakazaki, N., *et al*. (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis sp*. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109-136.
Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
Li, J., Brick, P., O'Hare, M. C., Skarzynski, T., Lloyd, L. F., Curry, V. A., Clark, I. M., Bigg, H. F., Hazleman, B. L., Cawston, T. E. & Blow, D. M. (1995). Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed β-propeller. *Structure,* **3**, 541-549.
Libson, A. M., Gittis, A. G., Collier, I. E., Marmer, B. L., Goldberg, G. I. & Lattman, E. E. (1995). Crystal structure of the haemopexin-like C-terminal domain of gelatinase A. *Nature Struct. Biol.* **2**, 938-942.
Matsushita, K., Shinagawa, E., Adachi, O. & Ameyama, M. (1989). Quinoprotein D-glucose dehydrogenase of the *Acinetobacter calcoaceticus* respiratory chain: membrane-bound and soluble forms are different molecular species. *Biochemistry,* **28**, 6276-6280.
Murzin, A. G. (1992). Structural principles for the propeller assembly of beta-sheets: the preference for seven-fold symmetry. *Proteins: Struct. Funct. Genet.* **14**, 191-201.
Nakai, K. & Kanehisa, M. (1991). Expert system for predicting localization sites in Gram-negative bacteria. *Proteins: Struct. Funct. Genet.* **11**, 95-110.
Neer, E. J. & Smith, T. F. (1996). G protein heterodimers: new structures propel new questions. *Cell,* **84**, 175-178.
Nicholls, A. & Honig, B. (1993). *GRASP1.1*, Columbia University, New York.
Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1-6.
Olsthoorn, A. J. J. & Duine, J. A. (1996). Production, characterization, and reconstitution of recombinant quinoprotein glucose dehydrogenase (soluble type; EC 1.1.99.17) apoenzyme of *Acinetobacter calcoaceticus*. *Arch. Biochem. Biophys.* **336**, 42-48.
Olsthoorn, A. J. J., Otsuki, T. & Duine, J. A. (1997). $Ca^{2+}$ and its substitutes have two different binding sites and roles in soluble, quinoprotein (pyrroloquinoline-quinone-containing) glucose dehydrogenase. *Eur. J. Biochem.* **247**, 659-665.
Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
Ramakrishnan, C. & Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformation. *Biophys. J.* **5**, 909-933.
Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallog. sect. A,* **42**, 140-149.
Renault, L., Nassar, N., Vetter, I., Becker, J., Klebe, C., Roth, M. & Wittinghofer, A. (1998). The 1.7 angstrom crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller. *Nature,* **392**, 97-101.
Rudenko, G., Bonten, E., Dazzo, A. & Hol, W. G. J. (1996). Structure determination of the human protective protein: twofold averaging reveals the three-dimensional structure of a domain which was entirely absent in the initial model. *Acta Crystallog. sect. D,* **52**, 923-936.
Schlunegger, M. P., Grütter, M. G., Streiff, M. B., Olsthoorn, A. J. & Duine, J. A. (1993). Crystallization and preliminary crystallographic investigations of the soluble glucose dehydrogenase from *Acinetobacter calcoaceticus*. *J. Mol. Biol.* **233**, 784-786.
Sondek, J., Bohm, A., Lambright, D. G., Hamm, H. E. & Sigler, P. B. (1996). Crystal structure of a $G_A$ protein βγ dimer at 2.1 Å resolution. *Nature,* **379**, 369-374.
van Kleef, M. A. & Duine, J. A. (1989). Factors relevant in bacterial pyrroloquinoline quinone production. *Appl. Environ. Microbiol.* **55**, 1209-1213.

Varghese, J. N. & Colman, P. M. (1991). Three-dimensional structure of the neuraminidase of influenza virus A/Tokyo/3/67 at 2.2 Å resolution. *J. Mol. Biol.* **221**, 473-486.

Varghese, J. N., Laver, W. G. & Colman, P. M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature,* **303**, 35-40.

Vellieux, F. M., Huitema, F., Groendijk, H., Kalk, K. H., Frank, Jzn J., Jongejan, J. A., Duine, J. A., Petratos, K., Drenth, J. & Hol, W. G. (1989). Structure of quinoprotein methylamine dehydrogenase at 2.25 Å resolution. *EMBO J.* **8**, 2171-2178.

Wall, M. A., Coleman, D. E., Lee, E., Iniguez-Lluhi, J. A., Posner, B. A., Gilman, A. G. & Sprang, S. R. (1995). The structure of the G protein heterotrimer $G_{i\alpha 1\beta 1\gamma 2}$. *Cell,* **83**, 1047-1058.

Xia, Z.-X., Dai, W.-W., Xiong, J.-P., Hao, Z. P., Davidson, V. L., White, S. & Mathews, F. S. (1992). The three-dimensional structures of methanol dehydrogenase from two methylotrophic bacteria at 2.6-Å resolution. *J. Biol. Chem.* **267**, 22289-22297.

Xia, Z.-X., Dai, W.-W., Zhang, Y.-E., White, S. A., Boyd, G. D. & Mathews, F. S. (1996). Determination of the gene sequence and the three-dimensional structure at 2.4 angstroms resolution of methanol dehydrogenase from *Methylophilus* W3A1. *J. Mol. Biol.* **259**, 480-501.

Ye, L., Hämmerle, M., Olsthoorn, A. J. J., Schuhmann, W., Schmidt, H.-L., Duine, J. A. & Heller, A. (1993). High current density ''wired'' quinoprotein glucose dehydrogenase electrode. *Anal. Chem.* **65**, 238-241.

***Edited by R. Huber***

*(Received 27 January 1999; received in revised form 31 March 1999; accepted 1 April 1999)*